

Avaliação de Desempenho

Tópicos Especiais em Arquitetura de Computadores

Quarto Trabalho Prático

Regras de Associação

Observações gerais sobre este trabalho

1. Este é um exercício de programação de análises de dados sobre bases de dados (datasets);
2. Você poderá programar em qualquer linguagem ou ambiente que você queira (Python, Shell script, R). Entretanto, a linguagem Python já tem muitas bibliotecas prontas (apyori, fpgrowth_py, mlxtend, sci-kit learn, numpy e pandas, por exemplo);
3. Se for aluno de **graduação**, poderá fazer este trabalho **individual**, ou em **duplas**; se for aluno de **pós-graduação** terá que fazer o trabalho **individual**;
4. Há diversos materiais disponíveis na Web sobre **regras de associação**, os quais podem ser consultados sem qualquer problema;
5. O livro texto dessa disciplina tem muitas dicas importantes e exemplos;
6. Você pode se basear no Jupyter Notebook disponibilizado nos materiais de estudos;
7. Considere a base de dados **Supermarket**.

Você deve escrever um **Notebook no Google Colaboratory** que resolva os problemas descritos abaixo:

- 1) Considere o dataset “supermarket.csv” disponibilizado nos materiais de estudo. A partir desse dataset escolha 40 (quarenta) itens dentre os 216 disponíveis no dataset **(1.0 ponto)**;
- 2) O dataset está no formato binário. Dependendo da biblioteca que você decida usar, pode se fazer necessário um **pré-processamento dos dados** de tal forma que todos os itens estejam no formato categórico;
- 3) Faça todos os procedimentos para gerar as principais regras de associação usando tanto o algoritmo Apriori quanto o algoritmo FP-Growth, para valores de **minsup** e **minconf** definidos por você **(4.0 pontos)**;
- 4) Para cada algoritmo, coloque as regras de associação retornadas em uma tabela de tal modo que fique mais fácil o acesso (serão duas tabelas). Cada linha da tabela deve conter o antecedente, o consequente, e os dados de suporte e confiança **(1.0 ponto)**;
- 5) Ordene as duas tabelas acima por suporte e confiança **(1.0 ponto)**;
- 6) Comentar o porquê dos valores definidos por você de **minsup** e **minconf** **(1.0 ponto)**;

7) Fazer comentários sobre os resultados das regras de associação e possíveis tomadas de decisão com relação aos itens que foram associados (exemplo: fazer promoção, colocar os itens próximos, venda casada, folhetos promocionais, etc). Justifique bem a tomada de decisão baseando-se nos fatos que os dados apontam **(1.0 ponto)**;

8) Comente sobre qual seria o ganho de adicionar mais métricas de desempenho, além da suporte e confiança? Quais seriam os ganhos? **(1.0 ponto)**